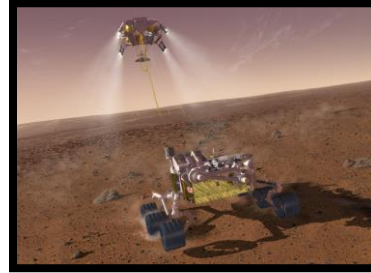




National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

219243



The PDS4 Information Model and its Role in Agile Science Data Curation

J. Steven Hughes and Daniel J. Crichton,

steve.hughes@jpl.nasa.gov, daniel.crichton@jpl.nasa.gov

NASA Jet Propulsion Laboratory (JPL)

California Institute of Technology

2017 American Geophysical Union Fall Meeting

New Orleans, LA December 11-15 2017

IN048. Planetary Data System: New Developments in Data Archiving

Tuesday December 12, 9:00 am



PDS4 Overview

- PDS4 is an information model-driven service architecture
 - *Supports the capture, management, distribution and integration of massive planetary science data captured in distributed data archives world-wide.*
- The PDS4 Information Model (IM) is the core element of the architecture
 - *Developed using lessons learned from 20 years of archiving Planetary Science Data*
 - *Used accepted standards for information model development*
 - Open Archival Information System (OAIS) Reference Model (ISO 14721)
 - Metadata Registry Specification (ISO/IEC 11179)
 - W3C XML (Extensible Markup Language) specifications.

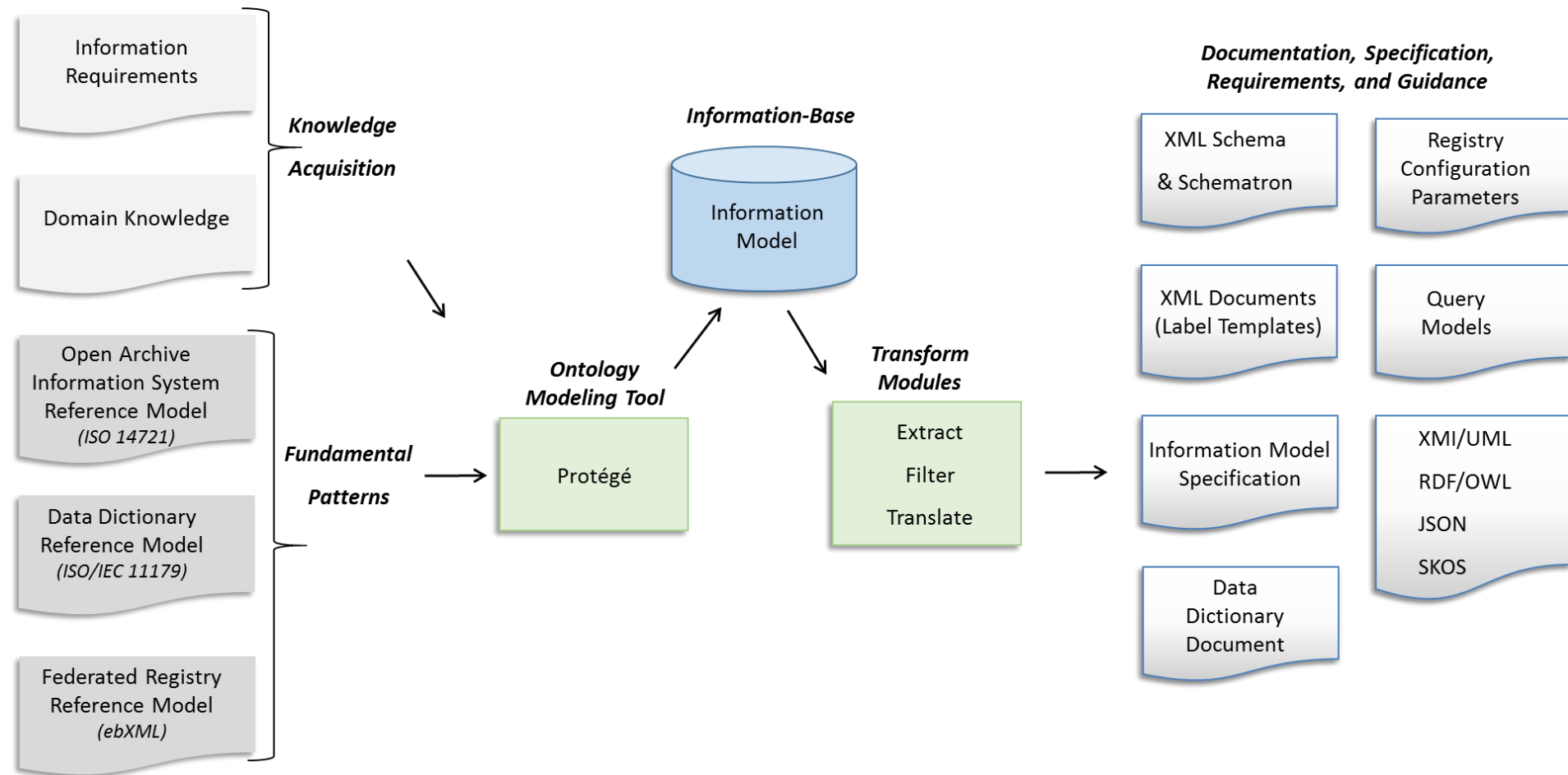


PDS4 Overview

- The PDS4 Information model is unique in that it drives the PDS4 infrastructure
 - *Provides the representation of concepts and their relationships, constraints, rules, and operations*
 - *Provides a sharable, stable, and organized set of information requirements*
 - *Provides machine parsable definitions that are suitable for configuring software and services and generating code.*



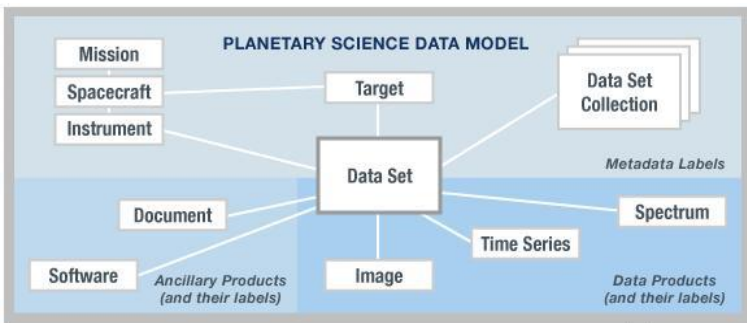
PDS4 Information Model (IM)



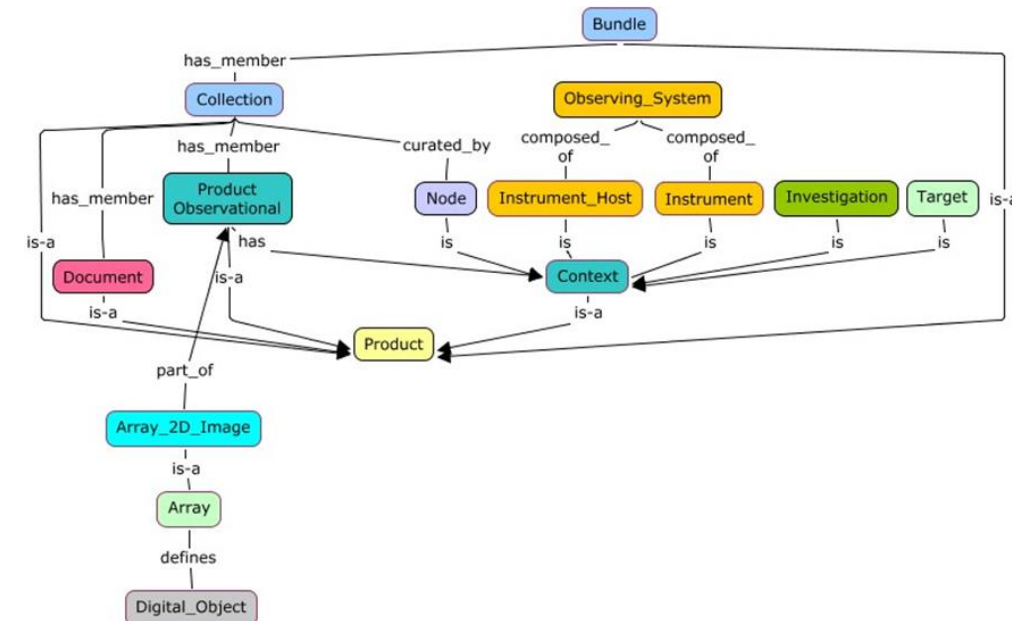


View Points

Community's View



Information Modeler's View



Repository View

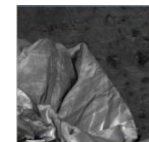
Product

Tagged Data Object

(Information Object)

```
<local_identifier>MPFL_M_IMP_IMAGE</local_identifier>
<offset unit="byte">0</offset>
<axes>2</axes>
<axis_index_order>Last_Index_Fastest</axis_index_order>
<encoding_type>Binary</encoding_type>
<Element_Array>
  <data_type>SignedMSB4</data_type>
  <unit>pixel</unit>
</Element_Array>
<Axis_Array>
  <axis_name>Line</axis_name>
  <elements>248</elements>
  <sequence_number>1</sequence_number>
</Axis_Array>
<Axis_Array>
  <axis_name>Sample</axis_name>
  <elements>256</elements>
  <sequence_number>2</sequence_number>
</Axis_Array>
</Array_2D_Image>
```

Describes



Data Object



PDS and Agile Data Curation

- PDS4 was developed with agile data curation¹ principles in mind.
 - *adaptive planning*
 - *early delivery*
 - *evolutionary development*
 - *continuous improvement*
 - *rapid and flexible response to change*
 - *maintain the value of the data over time*

¹ Young et al., Taking Another Look at the Data Management Life Cycle: Deconstruction, Agile, and Community, AGU 2014, Adopted from Agile Software Concepts, Manifesto for Agile Software Development, <http://agilemanifesto.org/>



Information Model Roles

- **Requirements:** The IM is the primary source for PDS information requirements.
- **Governance:** A multi-level governance scheme reduces the impact of change as the science community grows and evolves.
- **Configuration:** Extracts from the IM are used to configure tools and services
- **Semantics:** The IM provides named relationships to support semantic technologies
- **Usability:** PDS4 has been assessed that it is able to maintain the value of its data over time
- **Interoperability:** The PDS4 Standards have been adopted by the International Planetary Science Community



Requirements

- The PDS4 Information Model is the single authoritative source for writing PDS4 documents and standards
 - *Concepts Document*
 - *PDS4 Standards Reference*
 - *PDS4 Data Providers Handbook*
 - *XML Schema, Schematron, and XML documents for product label creation and validation*
- Initial drafts of the Information Model were available repeatedly during the development cycle for review, comment, and testing.
- Operational versions and point builds are available after Change Control Board approval and Integration and Test
 - *Builds currently occur every 6 months.*
- *Adaptive planning, Early delivery, Continuous improvement, Rapid and flexible response to change*

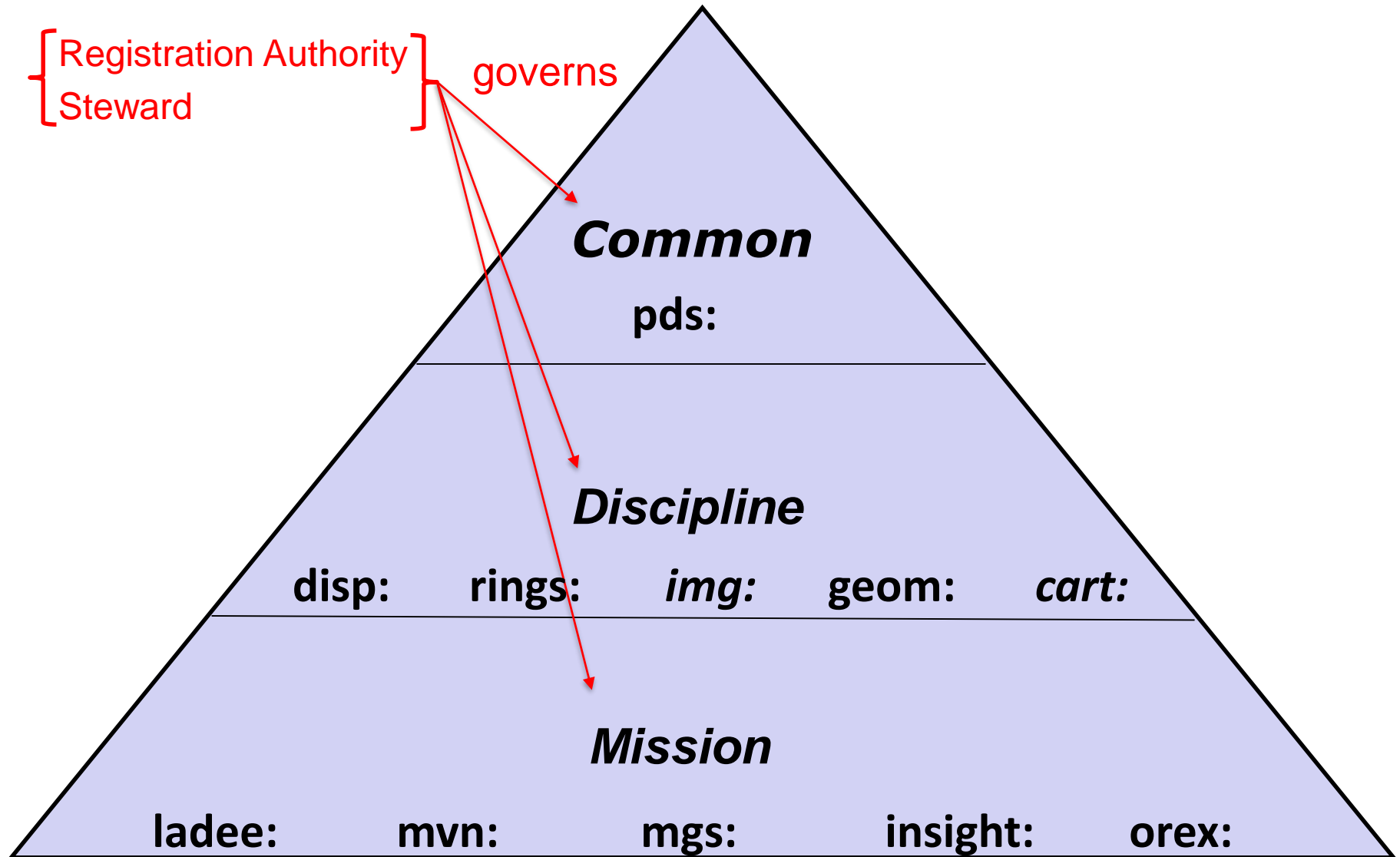


Multi-level Governance

- The PDS4 Information Model is partitioned into dictionaries (sub-models), each managed by a steward
 - *A single Common dictionary provides a stable foundation*
 - *Several Discipline dictionaries provide relatively stable "domains of discourse" for science disciplines within the community.*
 - *Many Project dictionaries provide local and relatively stable "domains of discourse" for each project/mission.*
- Evolutionary development
 - *The information components of the system can evolve independently and at a different speed than the technology components*



Multi-level Governance





Model Components

Common, Discipline and Mission Dictionaries

Registration Authority	Steward Id	Namespace Id*	XML Schema Namespace	Logical Identifier Prefix	Governance Level	Steward	Oversight
0001_NASA_PDS_1	pds	pds	http://pds.nasa.gov/pds4/pds/v1	urn:nasa:pds:	Common	PDS EN Node*****	CCB
0001_NASA_PDS_1	atm	atm	http://pds.nasa.gov/pds4/atm/v1	urn:nasa:pds:	Discipline	PDS ATM Node	
0001_JAXA_DARTS_1	darts	darts	http://pds.nasa.gov/pds4/darts/v1	urn:jaxa:darts:	Discipline	DARTS (JAXA)	
0001_NASA_PDS_1	en	dph	http://pds.nasa.gov/pds4/dph/v1	urn:nasa:pds:	Discipline	PDS EN Node	
0001_NASA_PDS_1	geo	geo	http://pds.nasa.gov/pds4/geo/v1	urn:nasa:pds:	Discipline	PDS GEO Node	
0001_NASA_PDS_1	geo	geom	http://pds.nasa.gov/pds4/geom/v1	urn:nasa:pds:	Discipline	PDS GEO Node	
0001_NASA_PDS_1	img	cart	http://pds.nasa.gov/pds4/cart/v1	urn:nasa:pds:	Discipline	PDS IMG Node	
0001_NASA_PDS_1	img	disp	http://pds.nasa.gov/pds4/disp/v1	urn:nasa:pds:	Discipline	PDS IMG Node	
0001_NASA_PDS_1	img	img	http://pds.nasa.gov/pds4/img/v1	urn:nasa:pds:	Discipline	PDS IMG Node	
0001_NASA_PDS_1	naif	naif	http://pds.nasa.gov/pds4/naif/v1	urn:nasa:pds:	Discipline	PDS NAIF Node	
0001_NASA_PDS_1	ops	pds	http://pds.nasa.gov/pds4/pds/v1	urn:nasa:pds:	Discipline	PDS EN Node	
0001_NASA_PDS_1	ppi	alt	http://pds.nasa.gov/pds4/alt/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_NASA_PDS_1	ppi	particle	http://pds.nasa.gov/pds4/particle/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_NASA_PDS_1	ppi	ppi	http://pds.nasa.gov/pds4/ppi/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_NASA_PDS_1	ppi	wave	http://pds.nasa.gov/pds4/wave/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_ESA_PSA_1	psa	psa	http://psa.esa.int/psa/v1	urn:psa:esa:	Discipline	ESA PSA	
0001_NASA_PDS_1	rings	rings	http://pds.nasa.gov/pds4/rings/v1	urn:nasa:pds:	Discipline	PDS Rings Node	
0001_NASA_PDS_1	rs	rs	http://pds.nasa.gov/pds4/rs/v1	urn:nasa:pds:	Discipline	PDS RS Node	
0001_ROS_RSSA_1	rssa	rssa	http://pds.nasa.gov/pds4/rssa/v1	urn:ros:rssa:	Discipline	RSSA (IKI)	
0001_NASA_PDS_1	sbn	sbn	http://pds.nasa.gov/pds4/sbn/v1	urn:nasa:pds:	Discipline	PDS SBN	
0001_NASA_PDS_1	sbn	sp	http://pds.nasa.gov/pds4/sp/v1	urn:nasa:pds:	Discipline	PDS SBN	
0001_NASA_PDS_1	atm	ladee	http://pds.nasa.gov/pds4/mission/ladee/v1	urn:nasa:pds:	Mission	PDS ATM Node	
0001_NASA_PDS_1	atm	ladee	http://pds.nasa.gov/pds4/ladee/v1	urn:nasa:pds:	Mission	PDS ATM Node	
0001_NASA_PDS_1	geo	insight	http://pds.nasa.gov/pds4/mission/insight/v1	urn:nasa:pds:	Mission	PDS GEO Node	
0001_NASA_PDS_1	img	mgs	http://pds.nasa.gov/pds4/mission/mgs/v1	urn:nasa:pds:	Mission	PDS IMG Node	
0001_NASA_PDS_1	img	mpf	http://pds.nasa.gov/pds4/mission/mpf/v1	urn:nasa:pds:	Mission	PDS IMG Node	
0001_NASA_PDS_1	sbn	orex	http://pds.nasa.gov/pds4/mission/orex/v1	urn:nasa:pds:	Mission	PDS SBN	
0001_NASA_PDS_1	ppi	mvn	http://pds.nasa.gov/pds4/mission/mvn/v1	urn:nasa:pds:	Mission	PDS PPI Node	
0001_NASA_PDS_1	ppi	mvn	http://pds.nasa.gov/pds4/mvn/v1	urn:nasa:pds:	Mission	PDS PPI Node	
0001_NASA_PDS_1	sbn	bopps	http://pds.nasa.gov/pds4/mission/bopps/v1	urn:nasa:pds:	Mission	PDS SBN	

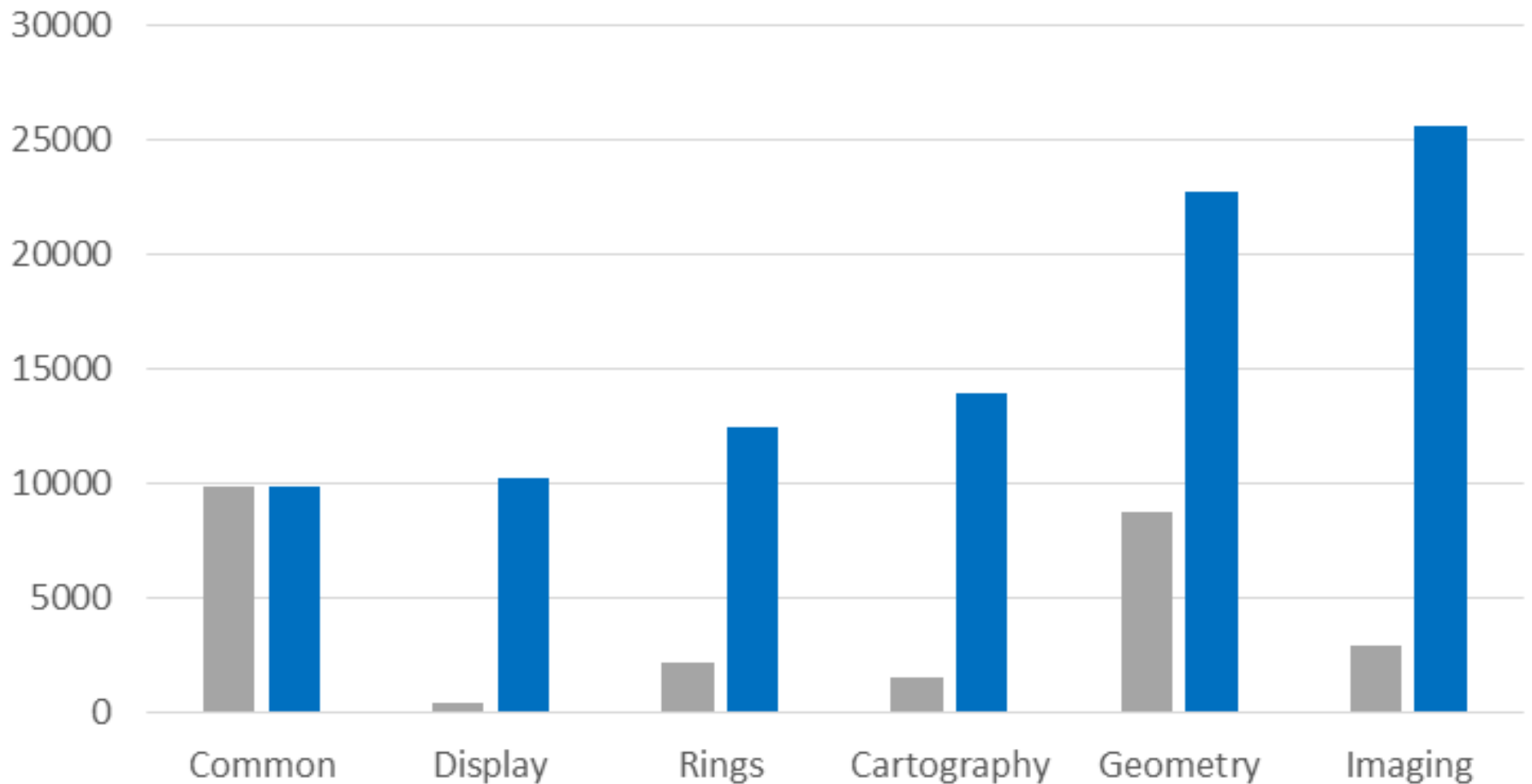


National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Common and Discipline Dictionaries

Information Requirements





System Configuration

- Software and services are designed to respond to the information model.
 - *Label Design*
 - *Label Generation*
 - *Label Validation*
 - *Registry configuration*
 - *Harvest and Search Query Models*
 - *Terminological mapping service*
 - *Data Analytics*
- *Rapid and flexible response to change*
- *Continuous improvement*



Semantic Technologies

- Under PDS4 all registry objects are first class products.
 - *All products have a Persistent Identifier (PID)*
 - *Products are used to describe data objects*
 - E.g., data, documents, people, software, and contextual objects.
 - *Product are inter-related with named relationships (semantics)*
- PDS has a pilot to obtain Digital Object Identifiers (DOIs) for its data collections.
- *Evolutionary development*
 - *Named relationships can be added to improve the semantic content*



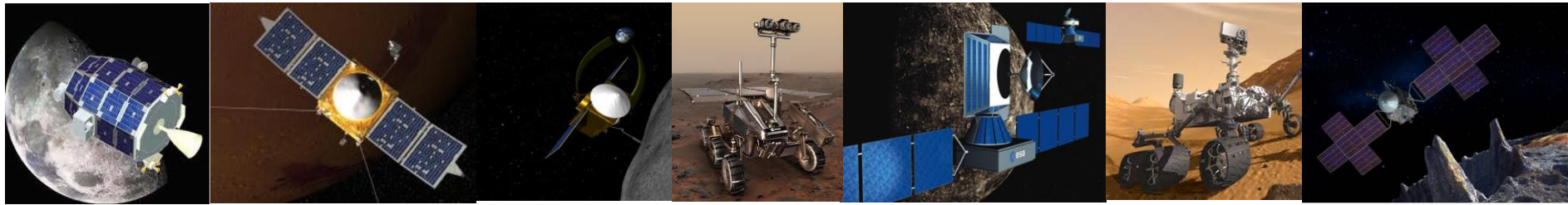
Maintain the Value of the Data Over Time

- A desk assessment of PDS4 against ISO 16363¹, the instrument for assessing a repository against the OAIS Reference Model², found that 92% of the metrics of the ISO 16363 standard were satisfied
 - 80% of the metrics for Governance and Organizational Viability
 - 95% of the metrics for Digital Object Management
 - 96% of the metrics for Infrastructure and Security Risk Management.
- Maintain the value of the data over time

¹ ISO 16363:2012 (CCSDS 652.0-R-1) Audit and certification of trustworthy digital repositories

² ISO 14721:2012 (CCSDS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model

International Collaboration on PDS4 Through IPDA



LADEE
(NASA)

MAVEN
(NASA)

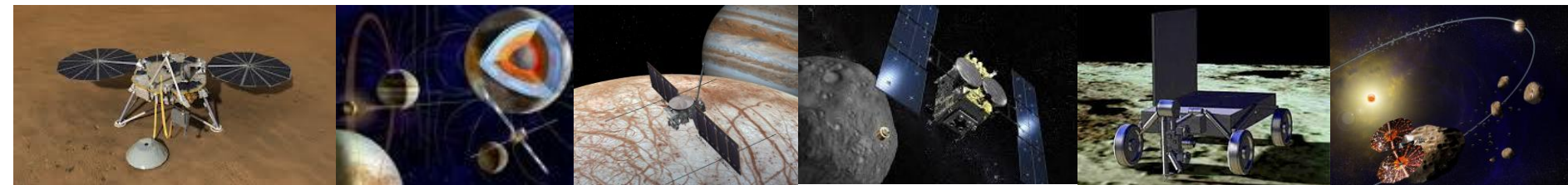
Osiris-Rex
(NASA)

ExoMars
(ESA/Russia)

BepiColombo
(ESA/JAXA)

Mars 2020
(NASA)

Psyche
(NASA)



InSight
(NASA)

JUICE
(ESA)

Europa
(NASA)

Hyabusa-2
(JAXA)

Chandrayaan-2
(ISRO)

Lucy
NASA



Summary

- PDS4 was developed with agile data curation principles in mind.
- The PDS4 IM provides a sharable, stable, and organized structure of information requirements
 - *adaptive planning*
 - *early delivery*
 - *evolutionary development*
 - *continuous improvement*
 - *rapid and flexible response to change*
 - *maintain the value of the data over time*
- The PDS4 Information Model and Standards are being used successfully across the International Planetary Science Community.



**National Aeronautics and
Space Administration**

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Thank You

Questions and Answers

PDS homepage: <https://pds.nasa.gov/>

Acknowledgements - This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

© 2017 California Institute of Technology. Government sponsorship acknowledged.



**National Aeronautics and
Space Administration**

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Backup



Usable Information

- Definitions for:
 - *data structure (format)*
 - *science interpretation of the data*
 - *context within which the data was captured, processed, and archived*
 - *relationships between the data*
- Expert knowledge from each of the science disciplines.
- Single authoritative source for the data standards.
- Drives the PDS4 infrastructure by providing:
 - *A sharable, stable, and organized structure of information requirements.*
 - *Formal definitions that are suitable for configuring and generating code.*
- Implementation agnostic



Information Model Overview

- Developed using:
 - lessons learned from 20 years worth of archiving
 - best practices for information model development
- Foundational principles adopted from:
 - Open Archival Information System (OAIS) Reference Model - ISO 14721 - Foundational Principles
 - ISO/IEC 11179 - Volume 3 – Metadata Registry Specification - Hierarchy of data dictionaries and governance model.
 - W3C XML (Extensible Markup Language) - Rules for encoding documents electronically.
- Drives the PDS4 infrastructure by providing:
 - the representation of concepts and their relationships, constraints, rules, and operations
 - a sharable, stable, and organized structure of information requirements.
 - formal definitions that are suitable for configuring and generating code.



Definitions

- “An information model is a representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse.”¹
- It provides a sharable, stable, and organized structure of **information requirements** or knowledge for the domain context.

¹ Lee, Y. T. 1999. Information Modeling: From Design To Implementation. In Proceedings of the Second World Manufacturing Congress, ed. S. Nahavandi and M. Saadat, 315-321. Canada/Switzerland: International Computer Science Conventions.